

Problems in the Science and Mathematics of ‘The Logic of Scientific Discovery’

Alan B. Whiting

School of Physics and Astronomy, University of Birmingham, United Kingdom. E-mail: abw@star.sr.bham.ac.uk

Editors: Ion C. Baianu & Danko Georgiev

Article history: Submitted on May 9, 2012; Accepted on August 4, 2012; Published on November 15, 2012.

Professor Sir Karl Popper (1902-1994) was one of the most influential philosophers of science of the twentieth century. However, in his most famous work *The Logic of Scientific Discovery* he displays troubling misunderstandings of science and mathematics at a basic level. These call into question his conclusions concerning the philosophy of science. *Quanta* 2012; 1: 13–18.

1 Introduction


Professor Sir Karl Popper was one of the most influential philosophers of science in the twentieth century. He is most widely known for his doctrine that scientific theories are not provable, but to be accepted as scientific they must be falsifiable. The most-cited statement of this is from the *Postscript* to *The Logic of Scientific Discovery* [1]:

... we adopt, as our criterion of demarcation, the *criterion of falsifiability*, i. e. of an (at least) unilateral or asymmetrical or *one-sided* decidability. According to this criterion, statements, or systems of statements, convey information about the empirical world only if they are ca-

pable of clashing with experience; or more precisely, only if they can be *systematically tested*, that is to say, if they can be subjected (in accordance with a ‘methodological decision’) to tests which *might* result in their refutation. [1, §*i, p. 315]

This was not a totally new idea. For instance, in 1902 Ernst Mach observed in passing, ‘Where neither confirmation nor refutation is possible, science is not concerned’ [2, p. 490], and citations of this type can be multiplied. Popper may be credited, however, with the emphasis on falsifying as opposed to confirming; and his popularity among practicing scientists is probably due to the straightforward way this criterion rejects much pseudoscience.

The book contains far more than this statement, however, including some truly troubling passages. Some of these I detail below. My aim here is not to present a complete evaluation of this book, for which a much larger and more detailed work would be necessary. Instead, I wish to highlight certain problems with Popper’s understanding of science and mathematics in *The Logic of Scientific Discovery*, without necessarily tracing their specific consequences.

 This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY-3.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2 The practice of science

Our first matter concerns *experimental uncertainty*. Experimental measurements are commonly accompanied by a statement of uncertainty or ‘experimental error’, as in 5.03 ± 0.05 in whatever units are appropriate. Roughly speaking, this means that the most probable value for the measurement in question is 5.03, but that values from 4.98 to 5.08 are quite possible. To be more rigorous: the *actual* value of which the measurement is an approximation is most probably 5.03; as one gets away from this figure, the probability decreases. The way it decreases depends in complicated way on many things, but in a surprising number of cases it is distributed in a Gaussian (normal) manner:

$$dP(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] dx \quad (1)$$

which gives (infinitesimal) probability dP of x having a value between x and $x + dx$. (The probability of finding exactly any given value, for a continuous variable, is zero; which is why the definition is given in terms of the infinitesimal range dx .) Here μ is the *mean* and σ is the *standard deviation* (Figure 1). If this is the case and 0.05 (in the expression 5.03 ± 0.05) is a single standard deviation, about 68% of the time the actual value is within the range indicated. If 0.05 is two standard deviations, more than 95% of the time the actual value will be within that range.

The convention of quoting one or more standard deviations is widespread, but not ubiquitous. In his 1906 textbook Newcomb [3, §29, pp. 53-57], for instance, begins his discussion of the matter using a 50% limit before proceeding to describe the normal distribution.

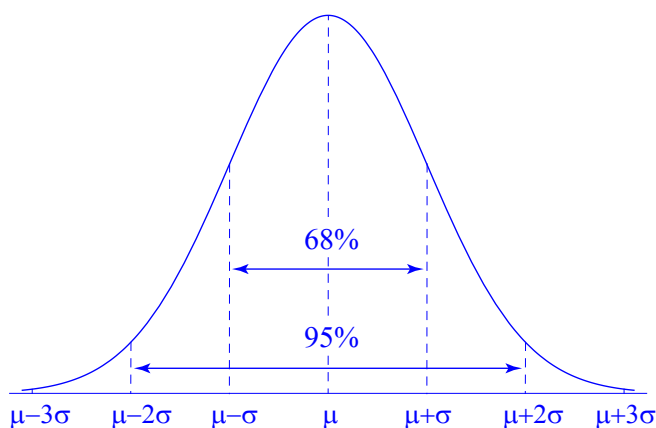


Figure 1: For a Gaussian (normal) distribution about 68% of the values lie within one standard deviation of the mean ($\mu \pm \sigma$), about 95% of the values lie within two standard deviations of the mean ($\mu \pm 2\sigma$) and about 99.7% of the values lie within three standard deviations of the mean ($\mu \pm 3\sigma$).

In no case can we state *certainly* that the actual value is within the limits, though we can get pretty close. Any competent scientist, in presenting a numerical result, will tell you whether the quoted uncertainty is one or two (or more) standard deviations, and whether the distribution is known to be, assumed to be, or known not to be, Gaussian. That is, the shape of the probable distribution around the quoted figure is Gaussian (or something else, as specified); the size is given by the \pm value.

Popper has a different interpretation. After asserting that every physical measurement is equivalent to noting a pointer’s position between two marks on a scale, which thus correspond to an interval within which the measurement lies, Popper continues:

It is the custom of physicists to estimate the interval for every measurement. (Thus following Millikan they give, for example, the elementary charge of the electron, measured in electrostatic units, as $e = 4.774 \cdot 10^{-10}$, adding that the range of imprecision is $\pm 0.005 \cdot 10^{-10}$.) But this raises a problem. What can be the purpose of replacing, as it were, one mark on a scale by *two*—to wit, the two bounds of the interval—when for each of these two bounds there must again arise the same question: what are the limits of accuracy for the bounds of the interval?

Giving the bounds of the interval is clearly useless unless these two bounds in turn can be fixed with a degree of precision greatly exceeding what we can hope to attain for the original measurement; fixed, that is, within their own intervals of imprecision which should thus be smaller, by several orders of magnitude, than the interval they determine for the value of the original measurement. In other words, the bounds of the interval are not sharp bounds but are really very small intervals, the bounds of which are in their turn still much smaller intervals, and so on. In this way we arrive at the idea of what may be called the ‘unsharp bounds’ or ‘condensation bounds’ of the interval. [1, §37, pp. 109-110]

By some unknown process, Popper has apparently interpreted the quoted uncertainties of a measurement as absolute limits; inside, the probability is unity, while outside it is zero. There are two points I wish to make about this passage. First, while such a naive misconception of experimental uncertainty might be understandable in someone who had never had contact with science at all, it is bizarre in a professor who is writing a book purporting to set out very basic aspects of science. Second, this

misunderstanding generates a whole conceptual structure (of ‘condensation bounds’), not only taking up space in itself but developing further ideas (notably [1, §68, pp. 190-197]).

3 Mathematics

Most of the book is taken up with an analysis of the mathematical theory of *probability*, including criticism of others’ formulations and a detailed presentation of Popper’s own construction. Before treating probability as such, I want to look at a couple of examples of Popper’s use of mathematics. I will first set out some background material for the benefit of those unfamiliar with this part of analysis.

We start with *infinite sequences*. These may be of points in a plane or in space, though with Popper we need only worry about points on a line, and indeed numbers on some limited section of the line: between 0 and 1, say. An infinite sequence is any infinite collection of them, taken one at a time in a certain order. Such a sequence may be the relatively boring 0.5, 0.5, 0.5, ... forever, or may jump around and never spend more time one place than another. An infinite sequence may *converge* to some value, which means that, for any particular distance δ you can specify, all the members of the sequence after some n th member are closer to the value than that distance.

(It may be that no member of the sequence you can name actually has the convergence value. This whole apparatus of ‘for any δ ... some n th’ may seem a devious and unnecessary way to proceed, but it’s the way to be rigorous when dealing with infinities and infinitesimals, and is quite necessary.)

A sequence has an *accumulation point* if, for any distance you can name (call it δ again), there is some point of the sequence which is not the accumulation point, closer than δ to the accumulation point. Only an infinite sequence can have an accumulation point. This is because of the Hausdorff property, which holds in that part of mathematics in which we will be working, and says (roughly) that between any two distinct points there is always some non-zero distance. So if you specify any point of the sequence and the accumulation point, you’ll need a later point of the sequence to put between them if someone decides to use a smaller δ . Any infinite, bounded sequence has at least one accumulation point, and may have more. (This material with some additions is found in [5, pp. 220-222])

Turning to Popper, he is concerned with the definition of a probability apparently given by von Mises. In a sequence of events, the fraction is formed of ‘successes’ in which a particular thing occurs divided by the total

number of events. If this fraction, called by Popper the ‘relative frequency,’ converges to a definite number in the limit of an infinitely long sequence, that number is the probability of success. (Popper does not quote von Mises explicitly. From the description it appears that the latter dealt with ‘convergence in probability,’ a phrase that Popper does not mention, but which he would not have cared for.) The particular example used is the fraction of ones in a sequence of ones and zeroes.

Popper wishes to do without the requirement of convergence. That means, he argues, he needs a concept that can be used in place of a limiting frequency, applicable to all infinite sequences.

One frequency concept fulfilling these conditions is the concept of a *point of accumulation of the sequence of relative frequencies*. (A value a is said to be a point of accumulation of a sequence if after any given element there are elements deviating from a by less than a given amount, however small.) That this concept is applicable without restriction to all infinite reference sequences may be seen from the fact that for every finite alternative *at least one* such point of accumulation must exist for the sequence of relative frequencies which corresponds to it. Since relative frequencies can never be greater than 1 nor less than 0, a sequence of them must be bounded by 1 and 0. And as an infinite bounded sequence, it must (according to the famous theorem of Bolzano and Weierstrass) have *at least one* point of accumulation. [1, §64, p. 176]

Popper’s definition of accumulation points is acceptable for our purposes, and his statement of the Bolzano-Weierstrass theorem is accurate, except for the application to a ‘finite alternative.’

The theorem only holds for infinite sequences; in fact, as noted, a finite sequence has no accumulation points. In the paragraph above, his mention of ‘finite sequences’ suggests (at least) that he has not grasped this requirement; this is proved on the next page [1, p. 177 footnote 4], where he applies the concept of accumulation points to finite and infinite sequences indiscriminantly. Let me be explicit: Popper is using a theorem that just isn’t true.

The real problem with using accumulation points is that, while the Bolzano-Weierstrass theorem assures us of at least one, there may be many. Popper realizes this [1, p. 177 footnote 2], and spends some time on the rather trivial point that, in this case, they are not useful in defining probability. He also recognises that to require a unique accumulation point is equivalent to requiring

convergence [1, p. 177]. Then he requires uniqueness anyway, asserting that it isn't, and that in any case he is free to choose such sections of any sequence as have the behaviour he desires [1, p. 177 footnote 4].

I will pass over these last two problems with mathematical logic, because in fact accumulation points are irrelevant to the task Popper attempts in this section. The importance of this episode is not so much that Popper makes mistakes in the handling of accumulation points, which could be considered a rather esoteric bit of analysis; nor even in his failure to distinguish between finite and infinite, though that is certainly a serious drawback for anyone trying to do mathematics. It lies in the fact that he is not competent in this whole area of analysis, deploying irrelevant machinery and doing that improperly.

(Further criticism of this section of Popper's work is possible, but the point has been made. Actually, the very use of sequences of this sort as a basis for a theory of probability is untenable. The proof is now given to undergraduate mathematicians as an exercise; see [6, pp. 25, 225])

Let us look at another section of mathematics, set theory. Popper refers to Kolmogorov's development of a theory of probability that explicitly uses sets. But he does not like it:

And yet, he [Kolmogorov] assumes that, in ' $p(a, b)$ '—I am using my own symbols, not his [that is, the probability of a given b]— a and b are *sets*; thereby excluding, among others, the logical interpretation according to which a and b are statements (or 'propositions', if you like). He says, rightly, 'what the members of this set represent is of no importance'; but this remark is not sufficient to establish the formal character of the theory at which he aims; for in some interpretations, a and b have no members, nor anything that might correspond to members. [1, §*iv, pp. 330-331]

Nowadays set theory is taught in elementary school; I am not sure what its status was when Popper wrote. But among the very first concepts one comes across is \emptyset , the empty set, the set with no members. Also among the first concepts is that a set may be made up of *anything*, including points, statements, propositions, truth-values, complex numbers, apples, oranges—or all at once. Popper here asserts that these features are not found in set theory, a sign of very basic misunderstanding. In a later section [1, §*iv, pp. 350-351] he demonstrates the 'superiority' of his 'Boolean' approach over the 'set-theoretic' approach *by performing set operations*. Somehow he has learned the label 'set theory' without noticing that it contains certain features he desires, and uses this flawed

understanding to attack Kolmogorov's formulation of probability.

In presenting these two examples I am not asserting that Popper's mistakes and misconceptions necessarily make all of his later work wrong. That would take a rather tedious effort of working through hundreds of pages of sometimes convoluted logic. I *am* asserting that he has attempted to produce results with mathematics that he does not understand or, worse, understands wrongly.

4 Probability

In what follows I will use Popper's notation as introduced above. That is, $p(a)$ denotes the probability of a , whatever a might be (heads, tails, the coin being unbiased, rain ruining our picnic). The probability of a given b , that is, we know that b occurred and now calculate the probability of a happening, is $p(a, b)$. Given both b and c , the probability of a becomes $p(a, bc)$. If we know that b did *not* occur, we have $p(a, \bar{b})$.

Popper takes issue with the idea that a theory, a hypothesis, may be assigned a probability based on a series of observations.

Let us now try to follow up the suggestion that the hypotheses themselves are sequences of statements. One way of interpreting it would be to take, as elements of the sequence, the various singular statements which can contradict, or agree with, the hypothesis. The probability of the hypothesis would then be determined by the truth-frequency of those among the statements which agree with it. But this would give the hypothesis a probability of $1/2$ if, on the average, it is refuted by every second singular statement of the sequence! [1, §80, p. 255]

After considering a few modifications of this idea, he concludes,

This seems to me to exhaust the possibilities of basing the concept of the probability of a hypothesis on that of the frequency of true statements (or the frequency of false ones), and thereby on the frequency theory of the probability of events. [1, §80, p. 258]

There are two points to make about these statements immediately. One is that Popper has set up an algorithm for hypothesis testing that would not for a moment be entertained by a competent scientist actually attempting to test a hypothesis. The second is that, in presenting this algorithm, he has set up a 'strawman,' any by refuting it

has pretended to refute *all* methods of testing a hypothesis (and giving it a greater or lesser probability) by examining events. (He may have derived this formulation from an equation he attributes to Jeffries, [1, §*vii, p. 383], which he does not appear to understand.)

Next I want to bring out two other examples presented by Popper concerning basic calculations in probability. First, he wants to disprove the ‘subjectivist theory of evidence,’ what we would now call the Bayesian approach [1, §*ix, pp. 425-426]. Given a coin, what is the probability of heads? Well, one-half. Now given that the same coin gave 500,000 heads (± 1350) in one million trials, what is the probability? One-half again. Hence, under the subjectivist theory, after a million coin flips *we have learned nothing*.

Popper believes this exercise disproves the idea of associating a probability with a subjective belief in an outcome (the initial guess of one-half), what we would now call a Bayesian prior. It is actually an exercise in asking the wrong question (a useful one would be: how certain are we that the coin is a fair one?). It also brings out Popper’s inability to distinguish, functionally, between a coin and a hypothesis.

Before we go on to the question of the absolute probability of a theory, there is one section that I think illuminates Popper’s ideas in an interesting way. In [1, §*ix, pp. 402-7] he is concerned with the question of how much a given event, a test in his words, will corroborate a theory. He wants to show that the standard theory of probability cannot calculate this, because in using it

... we should be forced to adopt a number of highly paradoxical views, among them the following clearly self-contradictory assertion:

(*)There are cases in which x is strongly supported by z and y is strongly undermined by z while, at the same time, x is confirmed by z to a lesser degree than is y . [1, §*ix, p. 405]

What he means by strong support is set out on the next page, using an example that deals with the throw of a die. We take $p(x)$ as the probability of throwing a six, $p(\bar{x})$ that of the negation (throwing anything else). Initially, with no information, we set the probabilities as 1/6 and 5/6. Then, we are given the information that that throw is an even number. The probabilities are now 1/3 and 2/3. The information has supported the hypothesis x and weakened its negation \bar{x} ; the probability of x is still smaller than \bar{x} . As it stands, this result is conceded to be ‘far from paradoxical.’ But if probabilities are interpreted as degrees of corroboration, it leads to the statement (p. 407) that a certain bit of information corroborates x and acts conversely on its negation, while x still has a lower degree of

confirmation than its negation; which Popper pronounces ‘clearly self-contradictory.’ He thus concludes that any calculation of the support or refutation of a hypothesis (a theory) *cannot* be done within conventional probability.

Popper is requiring a system of corroboration and confirmation in which *any* corroboration results in a sort of absolute confirmation. Where this idea comes from is not clear; Popper seems to consider it self-evident. And he is certainly at liberty to produce technical definitions of corroboration and confirmation in such a way, if he desires. But in asserting that because normal probability does not conform to this (essentially a bit of personal prejudice), *any* system calculating support or otherwise for a theory cannot be based on normal probability (p. 405), he certainly goes too far.

We now come to Popper’s calculation of the probability of a theory, *any* theory, in the universe. It is probably set out most clearly in [1, §*vii, pp. 374-380]. The probability of a theory is set equal to the product of the probabilities of all events predicted by the theory

$$p(a) = \prod_n p(a_n) \quad (2)$$

Since each component $p(a_n)$ can never be greater than unity, may in fact be less, and they will be infinite in number, the product will always go to zero. Hence *the probability of any theory is exactly zero*.

I have already noted that this is not the way to work out the probability of a hypothesis, given data. It also assumes that *all events in the universe are independent*. (This follows directly from his formula: the probability of a collection of events all happening is equal to the product of their individual probabilities only if they are independent ([5, p. 602]).) Popper formally recognises this, but asserts obscurely that any dependence of one event on another would be ‘non-logical.’ What the total independence of all events actually implies is total chaos—no event would have any relation to any other, and from one moment to another, from one point to another, *anything* could happen.

Popper explicitly concludes [1, §*vii, p. 375] something almost as depressing. If a is any theory and b is any information, *always*

$$p(a) = 0 \quad (3)$$

$$p(a, b) = 0 \quad (4)$$

Consequences of this result will be traced in the next section. For the moment, let me emphasize that Popper’s understanding of elementary probability was inadequate and flawed, not attaining the point of being able to test a possibly biased coin.

5 Popper's probability

While it is not my purpose here to trace the detailed consequences of Popper's flawed understanding, I suggest it is illuminating to look at one set of results.

Popper's conclusions led him to construct a 'calculus of relative probability' in which all of the following formulae may be valid [1, §*iv, p. 335]:

$$p(a, b\bar{b}) = 1 \quad (5)$$

$$\text{If } p(\bar{b}, b) \neq 0 \quad \text{then } p(a, b) = 1 \quad (6)$$

$$\text{If } p(a, \bar{a}b) \neq 0 \quad \text{then } p(a, b) = 1 \quad (7)$$

These allow some very strange things. In the first, some situation a is certain to happen, given both b and its negation. In the second and third, a situation is contemplated in which the simultaneous occurrence of something and its negation is given a nonzero probability. That is, the given situations would require a single flip of a coin to be both heads and tails.

In addition, one could also have simultaneously (for theories a_1 and a_2),

$$p(a_1, a_2) = 0 \quad (8)$$

$$p(a_2, a_1) = 1 \quad (9)$$

while at the same time $p(a_1) = p(a_2) = 0$ [1, §*vii, p. 388]. Eventually he introduces a notion to express the fact that, while the probability of every theory is zero, some zeroes are bigger than others [1, §*vii, pp. 388-389].

The kindest thing to say about this system is that it assigns meaning to something essentially meaningless. The first three lines could be fairly paraphrased in words starting with, 'If something that never happens, happens ...' which, if not nonsense, is very close to it.

He does not appear to notice the inconsistency of this with his claim to be implementing Boolean logic [1, §*iv, p. 332]. (Contemplating the possibility of both b and \bar{b} to happen is directly contradictory to the Boolean requirement that a statement be either true or false [5, p. 501].)

This system is quite possibly useless for any real science; it is certainly not used by scientists.

6 Implications

I must point out that, in criticising a philosopher's understanding of science and mathematics, I do not mean to imply that all philosophers must be highly competent scientists or mathematicians. It is quite possible to make serious contributions even to the philosophy of science without a specialist's deep background in the subjects;

although one must use care when working around one's ignorance.

That is not the situation we have here. Karl Popper's knowledge of certain parts (at least) of science and mathematics, at a basic level, was simply wrong. His understanding of the concept of experimental uncertainty was erroneous. His comprehension of classical probability was inadequate to test a possibly biased coin. He made errors in the mathematics of set theory and infinite series. The level of understanding of these fields displayed in his book would fail an undergraduate course in science and statistics.

These problems have consequences, though I have not attempted to trace them all here. His critique of standard probability theory is based on his own misunderstandings (in setting up the formula to test a hypothesis, in erroneous assertions about set theory), and include at least two requirements (that all events are independent, that any support for a theory must make it absolutely more probable than alternatives) that are at best arbitrary and are arguably simply wrong.

Concerning his strictly philosophical ideas and writings I have nothing to say; this study does not touch on them. But it is difficult to justify attention to statements about the logic of scientific discovery made by anyone with such a basic misunderstanding of the subject.

References

- [1] Popper KR. *The Logic of Scientific Discovery*. London, New York: Routledge, 2002 (original German edition published in 1934, in English 1958).
- [2] Mach E. *The Science of Mechanics*, second revised edition. McCormack TJ (translator), Chicago: Open Court, 1902.
- [3] Newcomb S. *A Compendium of Spherical Astronomy with its Applications to the Determination and Reduction of Positions of the Fixed Stars*. New York: MacMillan and Co., 1906.
- [4] Mermin ND. *Boojums All the Way Through: Communicating Science in a Prosaic Age*. Cambridge: Cambridge University Press, 1990.
- [5] Bronshtein IN, Semedyayev KA. *Handbook of Mathematics*, third edition, Hirsch KA (editor). Heidelberg: Springer-Verlag, 1998.
- [6] Williams D. *Probability with Martingales*. Cambridge: Cambridge University Press, 1991.